Electronic Field Book Processing Technical Documentation
Table of Contents

Table of Contents ((continued)

Topic page

Chapter 1. Introduction

1.1 Purpose of technical documentation

The Electronic Field Book Processor (EFBP) uses a wide variety of mathematical techniques in surveying and statistics. Every user does not need to fully understand the theoretical basis for these techniques.
Understanding what the report files indicate are more important. Most users can simply refer to the EFBP user's guide for most questions that deal with production issues. A user unfamiliar with EFBP would need to read the EFBP user's guide before taking advantage of the information in this document.

There are times when a user wants to look at some broader background descriptive information on certain algorithms in EFBP. This document serves that purpose, and points the to more source documents of information as this is not intended to eliminate those references.

In addition, many users are asked technical questions by others go beyond their knowledge base. Those questions can now be forwarded to this document.

This document should be used in conjunction with the EFBP User's Guide and information realted to your particular data collector system. A user not familiar with EFBP would need to refer to the EFBP user's guide before utilizing information in this document. The Electronic Field Book (EFB) was developed for field survey collection by the Florida Department of Transportation, and the subsequent processing of it by EFBP. While EFB relies exclusively on EFBP for coordinate production, EFBP accepts other field system's survey measurements if the data is translated to the ascii raw data file format (.obs) which is read by EFBP. The .obs file and its format must be understood (see EFBP user's guide) before understanding this documentation as many references are made to its contents.

1.2 Discussion of components of the documentation

Chapter 2 deals with correction of systematic errors in surveying measurements. Chapter 3 discusses analysis of repetitive survey measurements. Chapter 4 discusses how a weighted average can be used to derive a realistic "average" from repeated observations for the same measurement which have different error estimates. Chapter 5 discusses correction of systematic errors due to earth curvature and atmospheric refraction.

Chapter 6 and 7 discuss horizontal and vertical datums respectively.

Chapter 8 highlights state plane coordinate computations as they relate to EFBP and use of its generated coordinates in other software systems.

Chapter 9 discusses the automatic sideshot identification algorithms in EFBP.

Chapter 10 discusses the importance of error estimation in least squares analysis as it relates EFBP. Reasonable strategies which enable integration of various measurement types is presented.

1

Chapter 11 discusses how one validates the quality of survey measurements based on the least squares analysis output. Chapter 12 discusses the interpretation of least squares post-adjustment error estimates of final coordinates. Chapter 13 presents the basic theory of least squares analysis, its application to non-linear equations (most survey measurements are of this form), efficient strategies for solution, and generation of post-adjustment coordinate error estimates.

Chapter 14 is a glossary of EFBP terminology, and this is followed by references which further document the mathematical foundations of EFBP.

2

Chapter 2. Instrument calibration

All survey field instruments can contain systematic errors due to the nature of the mechanical components in them. Personal and environmental systematic errors can also exist. These errors can be minimized through proper field techniques and office processing mechanisms. Instrumental systematic errors can be minimized if a calibration process is performed,
and mechanical or mathematical means are used to correct any systematic error that is detected. EFBP uses mathematical means to correct systematic error in ways that are now discussed.

2.1 Systematic vs. random error

Error can be systematic or random. Systematic error follows some mathematical rules which can be modelled and corrected by proper techniques or survey data processing. Random error follows the laws of probability which are evaluated using statistical processes such as least squares analysis. Sources of error in surveying are instrumental, personal, or environmental.

Instrumental systematic errors in surveying can include total station/theodolite horizontal and vertical collimation errors, electronic distance/prism combined offset and scale errors, a differential level's line of sight not being horizontal, and a tape containing offset (short or long) or scale errors. Instrumental random errors are due to the mechanical nature of survey instruments being limited in absolute measuring ability.

An example of a personal systematic errors is not applying the correct pull to a tape. A good example of random personal error is our inability to point perfectly with a total station or theodolite.

Environmental systematic errors include earth curvature and atmospheric refraction. Heat waves, making pointing difficult, are an example of an environmental random error.

The distinction between systematic error and random error can become difficult in some cases. An electronic distance measuring (EDM) device is affected by temperature and pressure. At an instance of time there is a temperature and pressure that can be used to model systematic error corrections. You obviously would not record temperature and pressure every time you make a measurement, so it is difficult to define the drift in temperature and pressure as totally systematic or random.

2.2 Total station/theodolite

If the vertical circle of a theodolite was in perfect adjustment, zero degrees would be at the zenith. A horizontal circle would be in perfect adjustment (except for graduation errors in a theodolite) if the direct and reverse circle readings when pointing, with no personal error, to the same object would differ by exactly 180 degrees. Both of these errors are minimized by measuring equal number of times in direct and reverse position and averaging.

3

Many situations, such as topographic data collection, do not justify repeated measurements in direct and reverse. It is still highly desirable to eliminate instrument systematic error in all measurements. This is why EFBP is able to process what are called numerical calibration records.

A theodolite/total station calibration is an equal number of direct and reverse readings to the same object. Obviously the object should be very well defined so precise pointings can be made. While 1 direct and 1 reverse reading will suffice, multiple pointings are recommended so the surveyor can ensure no blunders exist, obtain a more reliable measurement through averaging, and obtain an estimate of the operator's pointing error.

EFBP averages the direct readings and computes standard deviations in a single observation for horizontal and vertical pointings. The same is performed in the reverse position. The standard deviations indicate if a blunder exists, and in absence of a blunder indicate pointing ability.

If the instrument was in perfect vertical adjustment, the sum of the average direct and reverse vertical circles would be 360 degrees. The difference from 360 degrees represents twice the error. As an example, assume the average direct and reverse vertical pointings were 90-00-30 and 270-00-20. The sum is 360-00-50, and indicates every zenith circle reading should have 25 seconds subtracted from it. A sum less than 360 degrees would require an addition of the error value to all zenith circle readings.

The amount that the average direct and reverse horizontal circles are from being 180 degrees different again is twice the error. In this case the sign of the correction will be opposite as applied to horizontal angles measured in the direct and reverse positions. As an example consider the average direct and reverse horizontal circle pointings in a calibration to be 190-00-10 and 10-00-30 respectively. Horizontal angles measured in direct would have +10 seconds added to them, while the correction added to reverse horizontal angles would be -10 seconds.

Calibration values are applied to all measurements after it until another calibration that contains numerical data is reached (a user can store a calibration without circle readings using EFB as it contains other pertinent information). If no calibration exists at the beginning of an .obs file the calibration corrections are zero until a calibration with numerical data is reached. If two or more distinct calibrations are in consecutive order in an .obs file, the last is always used.

Due to the calibration process, EFBP treats direct and reverse readings as unique measurements because they can be corrected for the systematic errors which were the major reason for measuring in direct and reverse. Thus four direct and four reverse measurements are corrected for systematic calibration errors, summed, and divided by eight to obtain an average. Many methods used to recommend averaging direct and reverse measurements in obtaining 4 values which are then averaged. This is no longer necessary due to the calibration record.

4

2.3 Differential level

A differential level is calibrated for line of sight not being horizontal usually by a peg test. This test is described in all basic surveying texts, and usually involves one backsight/foresight combination at midpoint (this corrects systematic error) to a backsight/foresight combination where the sight distances are not equal.

EFBP performs no systematic correction based on a peg test calibration in an .obs file as it is assumed the surveyor adjusted the cross hairs as a result of the test to create the horizontal line of sight.

2.4 Electronic distance measurement (FIXIT)

An EDM/prism combination can contain offset (constant) and scale (parts per million - ppm) systematic errors. The offset error remains constant for any measured length of line. The scale error grows (or shrinks) as a linear function of the length of the line.

An EDM/prism combination can be tested for correct distance measurement by one of three methods:

(1) Lay out a precisely measured distance with a steel tape and compare that value to what you measure with the EDM. This will not resolve scale error as a precisely taped distance has to be short in length.

(2) Set two collinear points A, B, and C. Measured distances AB plus BC should equal measured distance AC. The difference is the error in the EDM for those given line lengths. Since this test usually uses short lines the scale error is usually not measurable.

(3) An EDM calibration range is utilized which has a series of known distances which vary in length. The shortest distance determines the offset error, and the scale error is modelled by how the difference between the known and measured distance values vary for different lengths of lines.
Public domain programs are available for computation of base line measurements.

Some total stations or EDM's allow a user to dial in corrections for offset and scale error so that the distances in a data file (.obs) will be already corrected for any systematic error influence. If this is not performed, an auxiliary program to EFBP called FIXIT allows you to apply offset and scale corrections to measured EDM distances before you actually process your data with EFBP. FIXIT has allowed an accidental prism offset or incorrect temperature/pressure problem to be efficiently corrected.

2.5 Tape (FIXIT)

A tape is like an EDM and is suspect to offset and scale errors. A tape is usually laid out on a known baseline to obtain this information. Taping is usually a small part of an .obs file. With a text editor one should separate total station and taping measurements before performing FIXIT type corrections.

5

The text editor can then be used to again merge the total station and taping measurements into one .obs file again.

6

Chapter 3. Repetition error, maximum spread, same line/different setup comparison

Repeated measurements provide the surveyor with checks for blunders and a way of estimating the quality of one's measurements. Certain types of repeated measurements are more generic in testing for blunders. As an example, repeated measurements from the same instrument setup does not check if a user set up the total station or prisms over the station(s) properly. Another setup at that station, or a setup which measures a common line but in the opposite direction, is a better check as the quality of the instrument setups can be made.

3.1 Simple averaging, standard error in a single observation, and standard error in the mean

These values can be derived from repeated measurements at the same instrument setup. Note standard error and standard deviation are used interchangeably in this discussion.

Slope distances, zenith angles, height of instruments, and height of targets are converted to horizontal distance and mark-to-mark elevation change before any averaging or standard error computation begins. This is especially required for elevation differences in case a change in height of target occurred during the repetition process.

Horizontal circle readings are converted to horizontal angles based on a unique backsight station before averaging and standard error computations begin. This allows any movement in the horizontal circle plate to be accounted for between repetitions. This is analogous to the process of "moving" an initial horizontal circle reading when turning a series of repetitions. The difference between horizontal circle readings (horizontal angle) provides a more generic comparison mechanism.

The backsight station for a particular setup is the station which was sighted the most times. If a given number of stations were sighted the same number of times, the first station of that group after the setup record is selected as the backsight. The .obs file is usually time sequenced, and thus the backsight is commonly measured to first at a setup.

The EFBP users guide has a series of examples of averaging and standard error computations, and thus one can refer to these examples if one needs to look at numerical examples.

3.1.1 Simple averaging

A simple average is performed for any repeated measurements at a setup.
A simple average is the sum of the individual measurements divided by the number of repetitions. It does not take into account that the individual measurements could vary in quality. A weighted average, which is used by EFBP in some other

7

situations and is discussed later, can take into account measurements of varying quality.

One point to make is that the .obs file allows for multiple pointings to exist on the same position number and in the same (direct/reverse) face.
This is common when one has performed a large number of topographic measurements, and one wishes to "check in" on the backsight at the end of the setup. Some surveyors like to routinely check in on the backsight in some pre-defined chronological fashion. Slope distances/zenith angles of this fashion are treated as separate measurements. Thus 1 measurement in position 1 direct, 1 measurement in position 1 reverse, 3 measurements in position 2 direct, and 1 measurement in position 2 reverse of slope distance/zenith angle will be treated as 6 measurements.

The same is not true for horizontal circle readings as the 3 measurements in position 2 direct will be averaged. At this point a unique horizontal circle reading exists in position 1 direct, position 1 reverse, position 2 direct, and position 2 reverse. These four values are used in computing horizontal angles from horizontal circles, and the horizontal angles are then averaged. Thus one unique horizontal circle reading exists for each position number and face (direct/reverse) prior to the horizontal angle computation/averaging process.
This eliminates the problem of how many horizontal angles exist from station A1 to A2 in position 1 reverse if station A1 was measured to 4 times and A2 once.
By EFBP's algorithm the four circle readings to A1 are averaged and one angle for that position number and face is computed.

The averaging of horizontal circle readings on the same position/face produces the multiple pointing error value in the abstracting (.gen) report.

Reiterating, slope distances and zenith angles are not averaged, instead their reduced horizontal distances and elevation differences are averaged.
It is possible to average a horizontal distance(s) in .obs with those in .obs that are derived from slope distance/zenith angles. Horizontal circle readings on the same position and face are averaged before horizontal angle computation occurs. Each position/face produces a horizontal angle which is subjected to the averaging process.

The simple averaging process is thus a fairly simple computation which is well documented in statistics texts and surveying textbook sections on statistics as it applies to surveying.

3.1.2 Standard error in a single observation

The standard error is the square root of the variance. To compute a standard error one must first compute the difference between each individual observation and the average. These are residuals in a averaging process. The residuals are squared and summed. That sum is divided by the
number of observations minus one to obtain the variance. The square root of this variance is the standard error in a single observation.

8

Residuals are squared because they will be both positive and negative in sign - a simple sum of residuals from a simple average yields zero. The division by the number of observations minus one is similar to dividing by the number of observations in simple averaging. The "minus one" is with reference to the number of observations beyond what you minimally need (one observation does indeed determine a value for that measurement). This could also be referred to as the number of checks or number of degrees of freedom. There is no standard error unless you make at least two measurements.

The physical meaning of the standard error in a single observation is if you made one more observation under the same conditions with the same equipment you would be approximately 67% (one sigma) confident that you would fall within the range defined between the average minus the standard error to the average plus the standard error (average plus or minus the standard error).

If your worst residual (maximum spread in the .gen report) is more than the standard error that is not cause to be alarmed. Approximately 33% of your data will fall outside, so obviously we cannot be discarding that amount of data as surely not all of those are blunders. Using three times the standard error as a blunder detection device brings approximately 95 % confidence that the outlier should be discarded. The three sigma rule for blunder detection is quite common across disciplines who make measurements.

3.1.3 Standard error in the mean (average)

The standard error in the mean is the standard error in a single observation divided by the square root of the number of observations. A standard error in the mean is half the magnitude of the standard error in a single observation if the number of observations is four.

The physical meaning of the standard error in the mean is if you made one the same number of observations under the same conditions with the same equipment you would be approximately 67% (one sigma) confident that you would fall within the range defined between the average minus the standard error in the mean to the average plus the standard error in the mean (average plus or minus the standard error in the mean).

The standard error in the mean reflects that increasing reliability by repetition has diminishing returns as the number of repetitions grow.
To first halve standard error in a single observation you need 4 repetitions. To halve it again you need 16 repetitions, and to halve it yet again you need 64 repetitions

3.1.4 Significance of standard errors

The standard error in the mean reflects the uncertainty in the average as opposed to a single observation. Obviously under the same measuring conditions a horizontal angle observed eight times will usually have less uncertainty than a horizontal angle measured twice. Other factors can play a role in this determination. If the angle measured eight times had a 12 meter backsight and 6 meter foresight you would probably feel less certain about it than the angle

9

measured only twice where the backsight and foresight distances were more than 500 meters. The "same measuring conditions" rule does not apply in that comparison. Likewise a setup with very short sight distances may by coincidence have its horizontal angle measured twice and repeat perfectly. The perfect repetition does not reflect our surveyor's knowledge of the problems with short sight distances.

The standard error derived from repetition is thus an estimate of uncertainty, and not an absolute mechanism in surveying for assessing data quality. Standard errors are simply based on repetition, and do not account for errors such as instrument positioning over a point or the leveling of the instrument. A standard error can thus be used more effectively if we add surveyor insight into the overall model.

3.2 Maximum spread

The maximum spread is the largest deviation of any single observation from the mean. In other words, it is the largest residual (absolute value) derived from an averaging process. The maximum spread is the best indicator of a blunder in a repetition process, while the standard errors are the better indicators of data uncertainty.

3.3 Same line / different setup comparison

One of the best checks of horizontal distances and elevation differences is measuring the line on more than one setup. In a traverse mode the second setup is usually at the sighted station of the first setup, and generally results from a prism being measured to on the backsight. This procedure is common as averaging elevation changes measured in opposite directions on a line can help eliminate most systematic errors due to earth curvature and atmospheric refraction.

This check is better than standard errors and maximum spreads as it does check for blunders that cannot be checked by simple repetition. These blunders include failing to check the setup over a point, leveling of an instrument, and some types of station naming problems. The use of it in blunder detection is described in the EFBP user's guide.

10

Chapter 4. Weighted averaging

Any horizontal distance or elevation difference measured on more than one setup is averaged at the same time the comparison of the two values is made for blunder detection purposes. A simple average is could be used in this process, but instead to take advantage of the data uncertainty estimates (usually called error estimates) generated by EFBP a weighted average is used. Error estimates can be generated totally by user supplied constant and ppm input error values, or can be generated as a function of the standard error in the mean plus user defined additions which are defined by constant and ppm add-ons. The error constants and add-ons are defined in EFBP's opening setup menu. The choice of user defined error estimates or standard error plus add-ons is also defined in this initial menu.

The weighted averaging process will first be illustrated. This will be followed by a discussion of the two choices in error estimate generation.

4.1 Example

Assume the first measurement of a line is 100.00 with an error estimate of 0.01 and the second measurement is 100.04 with an error estimate of 0.02. The weighted average is computed by:

$$\frac{[\ 100.00 * (1/0.01) + 100.04 * (1/0.02)\ ]}{[\ (1/0.01) + (1/0.02)\ ]} = \frac{[\ 100.00 * 100 + 100.04 * 50]}{[100 + 50]}$$

which produces an average value of 100.013. The inverse of the error estimate becomes the weight, and thus the first measurement received four times the weight of the second measurement. This would be the same as a simple average of two 100.00 measurements with the 100.04 value. Note this allows the better measurement (smaller error estimate) to have more of an affect in the averaging process. This method is the same whether the measurement is a horizontal distance or an elevation difference.

The error estimate for the weighted average value will be

$$[\mathrm{sqrt}\ (2) * 0.01 * 0.02\ ] / [0.01 + 0.02] = 0.009$$

where sqrt (2) means the square root of 2.00. The formula for the error estimate of the weighted average is sqrt (2) * error est. #1 * error est. #2 divided by the sum of the error estimates.

4.2 Error estimation by standard errors plus add-ons

Assume a 300.000 meter slope distance and zenith angle were measured by repetition and produced standard errors in the mean for horizontal distance and elevation change of 0.002 m. and 0.008 m. respectively. Most surveyors were judge these values to be optimistic in being used directly as error estimates for reasons previously discussed. In EFBP's initial menu horizontal distance constant and add-on parameters of 0.005 m. and 10 ppm were entered respectively.

11

For elevation changes by trigonometric leveling constant and ppm errors of 0.007 m. and 100 ppm were entered.

The horizontal error estimate would be 0.002 + 0.005 +(10/1000000)*300 or 0.002 + 0.005 + 0.003 = 0.010 m. The elevation change error estimate would be 0.008 + 0.007 + (100/1000000)*300 = 0.008 + 0.007 + 0.030 = 0.045 m. The error constant add-on could be thought of accounting for setup error that affects linear measurements, and the ppm add-on is due to the fact that longer lines generally contain more error in these types of measurements. The ppm add-on is usually larger for trigonometric leveling than horizontal distance as common surveying practice acknowledges that error in trigonometric leveling grows faster for a longer distance than for horizontal distance.

With non-zero constant and ppm error add-ons even perfect repetitions, producing a standard error in the mean of zero, will receive a non-zero error estimate. This is important because perfect repetition definitely does not reflect that a measurement contains no error.

4.3 Error estimation by user definition

Some users may feel error estimation partially based on standard errors from repetition is not a desired procedure. The other option is totally by user definition independent of repetition error. EFBP allows user defined values for constant error for horizontal distance, constant error for elevation difference, and a single ppm error for both types of measurements. Assume a user has input horizontal constant error of 0.008 m., elevation difference constant error of 0.02 m., and a ppm error of 10 ppm. A 300 m. distance, independent of repetition error, would produce a horizontal distance error estimate of 0.008 + (10/1000000)*300 = 0.011 m. and an elevation difference error estimate of 0.02 + (10/1000000)*300 = 0.023 m. Logically, the constant error for trigonometric leveling should be larger than the constant error for horizontal distance as it is more difficult to measure.

If error estimates from repetition is not utilized, the weighted average turns into a simple average because the error estimate for all measurements of a line will be equal.

Even if error estimates from repetition is being used, there will obviously be situations where a measurement is not repeated at a setup. The error estimate will thus be generated from the error estimates (not add-ons) by user definition. It is thus possible to perform a weighted average of a value with an error estimate generated from standard error and add-ons with a value that was not repeated and thus has its error estimate derived from user input constant and ppm values.

4.4 How many distances and elevation differences of a line are in the least squares analysis?

Since multiple setups which measure a horizontal distance or elevation change are subjected to weighted averaging only one final averaged value is subjected to the least squares analysis. This helps ensure a proper number of degrees of

12

freedom which is based on the geometry of the survey network, and not enhanced by repetitive measurements of the same line.

4.5 Why angle weighted averaging does not occur in EFBP

Horizontal angles are not subjected to a weighted average for several reasons. The first is that horizontal angles on not as often measured multiple times on different setups - measuring to a prism on a backsight creates the need for weighted averaging of horizontal distances and elevation differences. A common field practice is to measure both the interior and exterior angles when traversing. If measured on separate setups, including both in the least squares has some validity as residuals can indicate a better fit of one of the angles, which in turn indicates a possible setup error in the angle with the larger residual. Multiple occupations of the same setup can create a multitude of angles with different backsights which in turn makes angle averaging difficult as it would require definition of a single backsight from all of these angles. This may not be how the field data was collected. It was thus decided that the complications of angle averaging made placing all of them in the least squares analysis the suitable solution.

13

Chapter 5. Earth curvature & atmospheric refraction correction

Elevation differences derived from a total station need to be corrected for earth curvature and atmospheric refraction. It is also possible to correct differential leveling for these values but due to short sight distances it is not being performed in EFBP as it is negligible.

Atmospheric refraction bends a line of sight downward and thus it causes a line of sight to be lower than if no atmosphere existed and the line of sight was not bent.

To consider earth curvature better, pretend atmospheric refraction does not exist. It would be desired for a line of sight to parallel the curve of the earth. This is not possible as the line of sight is straight, and thus this sight is above the curved line, which means earth curvature causes a zenith angle to be above where it should be.

Since the errors are in opposite directions, if they were equal in magnitude they would cancel. For a standard atmosphere, atmospheric refraction has only 1/7th the effect of earth curvature. Thus the cumulative affect is the line of sight is too high which creates a elevation change which is too negative, and thus a positive correction is always applied to the elevation difference for earth curvature and atmospheric refraction.

The amount of correction in feet is $0.0206*(F/1000)^2$ where F is the slope distance in feet. If in meters the correction is $0.0675*(M/1000)^2$ where M is the slope distance in meters.

Note the correction grows as a squared function of the distance. A sight distance of 100 ft. produces a correction of only 0.0002 ft. so it is insignificant. 500 and 1000 ft. sight distance produces corrections of 0.005 ft. and 0.021 ft. respectively. Since a trigonometric elevation difference derived from a 1000 ft. sight distance is rarely accurate to 0.02 ft., it could be stated the earth curvature and atmospheric refraction corrections are insignificant for normal surveying practice.

The user has the ability to toggle the correction on or off in the initial EFBP menu. Some total stations have an ability to also toggle the correction on or off. Be very careful as the correction is often only applied to reduced values, and raw data is stored in the .obs file.
To determine if a total station corrects raw data call your local dealer or measure to a precise locatable point with the correction on and off and see if any change if raw data is noted.

14

Chapter 6. Horizontal datums

EFBP can automatically reduce data to state plane coordinate projections in NAD 27 or NAD 83. It is also possible for someone to create their own user defined projection coordinate system which is not necessarily not sea level. A zone identification number of 9999 indicates a user defined zone.

Note horizontal datum and state plane zone (NGS zone number) are stored in the control .ctl file. A lack of horizontal datum and state plane zone indicates assumed coordinates are being used.

6.1 Assumed horizontal datum (no geodetic datum)

There are situations where a surveyor chooses to use an assumed coordinate system and apply no geodetic/state plane reductions. This is very useful for checking for blunders in survey measurements, and determining relative distance and bearing changes between stations. Use of state plane coordinates in a control file with no datum or state plane zone designation will produce incorrect state plane coordinates as no scale or elevation factors can be applied. It is thus suggested when using assumed coordinates to make them "look" very different than state plane coordinate values in that general area.

If you are using EFBP without a control file, the first setup is assigned horizontal coordinates of 10000,10000 and due north is assumed to the first station sighted at that setup.

6.2 North American Datum of 1927 (NAD 27)

This is based on the Clarke ellipsoid of 1866. The units for distance and state plane coordinates were defined in U.S. Survey Feet, and thus EFBP will only process in this datum in English units. This datum was created by fixing a latitude/longitude at station Meade's Ranch, Kansas, and fixing a geodetic azimuth to a nearby azimuth mark. The type of measurements which made up the geodetic control network for this datum was primarily triangulation as EDM's had not been invented. The production of coordinates had to occur without the use of computers! All geodetic and state plane coordinate production prior to 1986 was with respect to this datum.

6.3 North American Datum of 1983 (NAD 83)

This is based on the World Geodetic Reference ellipsoid of 1984. The units for distance and state plane coordinates were defined in meters, with conversion to U.S. survey feet or international feet left up to the user's
preference. No fixed control existed. The type of measurements now included traverse, doppler, and the global positioning system (GPS) in addition to triangulation. It resulted in a least squares adjustment of approximately 250000 stations and resulted in new (better) coordinates for stations which had coordinates in NAD 83. Not all stations which had NAD 27 coordinates were part of this adjustment, and it is thus often desirable to convert these coordinates to NAD 83. The National Geodetic Survey (NGS) has produced a public domain program called NADCON for this purpose (geodetic coordinates only). The Army

15

Corps of Engineers updated NADCON with more options such as NAD 27 state plane to NAD 83 state plane. This public domain program is called CORPSCON.

The exact conversion between meters and U.S. survey feet is 1 meter = 39.37 inches. The exact conversion between meters and international feet is 25.4 millimeters = 1 inch. If performing a survey in English units in NAD 83 the surveyor must know if the particular state he or she is in has passed legislation stating which foot should be used. If the state has not passed legislation one should find which foot is being used by the agency one is doing work for.

One should not worry if which foot is used for one's measurements. The difference in a 1000 ft. distance is only 0.002 ft., and thus not within the measuring ability of conventional survey measurements.

6.4 Other - region, local supernetwork, High Accuracy Regional Network (HARN), etc.

Due to the advent of the global positioning system several states and regions have found it desirable to create a high precision network of GPS observations, and perform a least squares analysis of it for coordinate production. This will use the NAD 83 ellipsoid and NAD 83 state plane zone constants. A station which has coordinates in NAD 83 and the high precision network will not be equal, and differences are usually less than one foot. This makes unlabelled coordinates nearly impossible to detect as NAD 83 or supernetwork. Supernetworks are usually labelled such as NAD 83 (90) which implies the supernetwork coordinates which were published in 1990.

EFBP permits tagging of any two digit year to a horizontal datum in the control (.ctl) file which will be also be placed in the final coordinate (.XYZ) file. A year greater than 82 indicates NAD 83 datum and state plane zone constants will be used. A year greater than 83 indicates the coordinates are referenced to a regional supernetwork.

No mixing of control coordinates from different datums in one job should ever occur as there are systematic shifts between them.

16

Chapter 7. Vertical Datums

A vertical datum is not significant numerically to EFBP as geodetic reductions such as scale factors, elevation factors, and convergence angles are only applied to 2-D (horizontal) measurements. Nonetheless, it is very important to label control (.ctl) and final (.xyz and .soe) elevations with a vertical datum number. A lack of vertical datum number indicates an assumed vertical datum is being used. A vertical datum is designated with a 2 digit number in the .ctl, .xyz, and .soe files.

7.1 Assumed vertical datum (no geodetic reference)

This would indicate the benchmarks used are with respect to some arbitrary reference. If you are using EFBP without a control file, the first setup is assigned an arbitrary elevation of 500.00 .

If you have a geodetic horizontal datum and an assumed vertical datum, the elevations should at least be derived from interpolating from a map with reference to a vertical datum. This is because elevation factors in the geodetic reductions are based on the elevations in the .ctl file, and it is assumed these elevations are with respect to a vertical datum.

7.2 National Geodetic Vertical Datum of 1929 - NGVD 29

This was the only national geodetic vertical reference until approximately 1993. It was the production of elevations from differential leveling which was compiled by NGS at that time. Elevations were published in feet, and a series of benchmarks near the coastline were held fixed to force the datum to be referenced close to mean sea level. All benchmarks and contour maps published prior to 1993 by NGS, the U.S. Geological Survey and other federal and state mapping agencies were with respect to this datum.

7.3 North American Vertical Datum of 1988 - NAVD 88

The plethora of leveling observations which succeeded NGVD 27 plus better gravity measurements created the need for a redefinition of the vertical datum in North America. While labeled NAVD 88, the elevations were not published until 1993. All elevations are published in meters, and a user follows the same logic as horizontal coordinates in converting to either U.S. survey or international feet. Only one benchmark (near the mouth of the St. Lawrence River) was held fixed in the least squares adjustment of more than 200000 benchmarks.

Elevations for the same benchmark in NGVD 29 and NAVD 88 will not be equal. Many benchmarks with NGVD 29 elevations were not included in the NAVD 88 and thus need translation to it. NGS has provided public domain program VERTCON for that purpose.

7.4 Local datum

17

It is possible in an area to have a local datum which is offset from either NGVD 29 or NAVD 88. This should be labeled in the .ctl and .xyz files with a vertical number other than 29 or 88.

18

Chapter 8. State plane projections

State plane coordinates are based on two types of projection systems - a Lambert Conic Conformal or a Transverse Mercator. States that are elongated north-south tend to use Mercator zones and states elongated east-west tend to use Lambert zones. Florida is an example of a state elongated in portions of the state in different directions, and is thus made up of both Lambert and Mercator zones. All zones have central meridians with defined longitudes which point true north-south except for one Mercator zone in Alaska where the central meridian is offset 45 degrees from true north. This is called an oblique Mercator projection.

In NAD 1927 state plane zones the size of state plane zones were limited by the fact that the difference between a grid distance and a ground distance reduced to the ellipsoid would not exceed 1/10000. The difference between these two distances is known as the scale factor. The scale factor varies according to your location in a zone, and deviates furthest from unity at the center and extremes (E-W in Mercator, N-S in Lambert) of the zone. Thus larger states have more zones than smaller states.

In NAD 83 some states decided to eliminate some zones which in some cases now makes the difference between grid and ellipsoid distance greater than 1/10000. Some states also changed some zone origins, central meridian longitude, or meridian lines of scale factor of one.

8.1 Lambert conical projection

The Lambert projection is a cone which intersects the ellipsoid at two defined longitudes where scale factor would be one. The scale factor does not change in an east-west direction.

8.2 Transverse Mercator projection

The Mercator is a cylindrical projection where the centerline of the cylinder is running in an east-west direction. The cylinder intersects the ellipsoid at defined longitudes.

8.3 Zone origin, false northings, and false eastings

An origin for the zone is defined by latitude and longitude. A false easting is assigned to the central meridian which prevented negative eastings. While sometimes the origin received a false northing, it was more common to set the false northing of the origin to zero as it was far enough south of the location of the zone to prevent creation of negative northings.

8.4 NAD 27 vs. NAD 83

To force NAD 83 state plane coordinates to look different than their NAD 27 equivalents two items were instituted. NAD 83 state plane coordinates were published by NGS in meters, while NAD 27 state plane coordinates were published in feet. In addition the false easting (and in some cases also the false

19

northing) were changed so that even if coordinates in NAD 83 were converted to feet they would not match their NAD 27 counterparts. In most cases false eastings in NAD 27 Lambert zones were 2000000 ft. in NAD 27 to and NAD 27 Mercator false eastings were 500000 ft. The false eastings in NAD 83 actually vary from state to state.

8.5 Difference between grid distance and ground distance

The grid distance between two points is simply the pythagoreum plane of the end point coordinates.

Grid distance is computed from ground (horizontal) distance by:

Grid distance = Ground distance * scale factor * elevation factor

and thus ground distance is computed from grid distance by:

Ground distance = Grid distance / (scale factor * elevation factor)

Scale factor for a line is usually computed by averaging the scale factors at the end points of the line. Remember the scale factor is a function of your location in the zone. The elevation factor is derived from the average of the end point elevations (ave. elev.) and in feet is computed by:

elevation factor = 20906000 / (20906000 + ave. elev.)

where 20906000 ft. is a suitable approximation for the radius of the earth. The metric equivalent of 20906000 can obviously be computed and then ave. elev. can be entered in meters.

It should be noted that it is theoretically correct to reduce to the ellipsoid, and not the geoid (elevation reference). The difference between the geoid and ellipsoid is approximately 20-30 meters in North America, which causes an error in elevation factor of approximately 1/200000. This makes it smaller than our usual random errors in surveying, and thus using elevation, not ellipsoid height, is valid.

8.6 Convergence angles and T-t corrections

A convergence angle is the difference between grid north and forward geodetic north at a point. A forward geodetic azimuth at a station is the angle from geodetic north to another station. Geodetic north lines converge to the north pole and therefore only parallel at the equator. Grid north (state plane north) lines are parallel to one another. The convergence angle is zero at the central meridian because grid north and geodetic north coincide.

The equation which relates azimuths in the two systems is:

grid azimuth = geodetic azimuth - convergence angle + T-t correction

or

20

geodetic azimuth = grid azimuth + convergence angle - T-t correction

Convergence angles are thus negative when west of the central meridian and positive when east of the central meridian. The T-t (second term) correction is insignificant on typical survey distances but can become a few seconds for lines longer than a mile near the edge of a zone.

Horizontal angles are reduced to grid by only T-t correction:

grid angle = geodetic angle + foresight T-t corr. - backsight T-t corr.

Again the T-t correction rarely exceeds tenths of seconds. This correction is due to the fact that horizontal angles measured on a curved earth need to be reduced to the flat state plane grid.

8.7 Forward vs. mean vs. reverse geodetic and astronomic azimuths

While different in format, the term azimuth equally applies to bearings in this discussion. In the previous section we have defined the relationship between grid and forward geodetic azimuths. A reverse (back) geodetic azimuth is the forward geodetic azimuth from the sighted station back to the occupied station. The forward and reverse azimuth do not differ by 180 degrees, except on a north-south line, because of convergence of meridians towards the north pole.

The mean geodetic azimuth of a line is the average of the geodetic forward and reverse bearings. It is a line of constant bearing and is thus a curved line on the face of the earth. An east-west section is an excellent example of a line that represents mean bearing as it is intended to be a line of constant latitude.

Astronomic azimuths are similar in nature to geodetic equivalents, except that its reference is astronomic north. Astronomic north is determined from surveying measurements to stars or the sun. The difference between astronomic and geodetic north is a function of the direction of gravity, and thus varies according to your location. In most parts of the United States the difference between astronomic and geodetic north is less than one second. NGS has a public domain program available called DEFLECT90 which outputs the difference between geodetic and astronomic north based on input latitude/longitude.

8.8 How does EFBP do state plane reductions?

The abstracting initial phase of EFBP identifies redundant stations, and generates preliminary coordinates for the least squares analysis by automatic coordinate geometry computations. Even without using state plane reductions, these preliminary coordinates are rarely most than 10 feet from their least squares adjusted values.

21

The second component of EFBP is the 1D least squares analysis and sideshot computations for all 1-D sideshots. This produces elevations for all stations which can be used for elevation factors for any 2-D coordinate computations.

The third component of EFBP is the 2D least squares analysis. It uses the preliminary coordinates from the abstracting stage to obtain point scale factors. The end point scale factors of a distance are averaged to obtain a scale factor for that line. Scale factors change minimally over survey measurement type distances, and thus the preliminary coordinates are as good as the final least squares adjusted values for scale factor generation. Every point has an elevation from the 1D computations and thus an average of the end point elevations for a line can be used to generate elevation factors.

If geodetic azimuths exist in the .ctl file, the preliminary coordinates are used to compute a convergence angle for reduction of that azimuth to grid. Similarly the preliminary coordinates are used to generate T-t corrections for all geodetic azimuths and horizontal angles.

Sideshots are generally fairly short lines and thus the scale factor change, elevation factor change, and T-t corrections will be negligible. Thus the horizontal sideshots, which are based on the least squares adjusted coordinates of the redundant stations, utilize the sideshot's occupied station's point scale factor and elevation factor.

Thus it has been defined how all measurements reductions to grid are automatically employed. It has also been defined how all sideshot computations are based on the results of the least squares analysis.

22

Chapter 9. Sideshot identification algorithm

EFBP requires no identification of sideshot vs. redundant observation or station. EFBP automatically identifies the sideshots via the "connectivity" of the survey as defined by observation station name.

Note a 1D sideshot can be a 2D redundant point, and vice-versa. A benchmark which is only measured from one station is a 2D sideshot, but is definitely not a sideshot vertically. A 2D (horizontal) control point that is only measured from one other station will not be a redundant 1D station.

Performing a 1D/2D analysis allows for the uniqueness that many times the redundant 1D network is different from the redundant 2D network because not all survey control is usually 3D in nature. The 1D/2D approach also allows one to integrate differential leveling, station-offset, 2-D traverse, and 3-D traverse into the same job. The 1D/2D approach has also been shown by the author (see references) to be more suitable for reduction of conventional survey measurements, and in all cases producing statistically the same results as a full 3D approach.

9.1 1D sideshot identification

The only measurements in the 1D analysis are elevation differences and benchmarks. An elevation difference connects two stations.

The 1D sideshots algorithm looks for stations that are not benchmarks that are only connected to one other station. These are sideshots and are "pruned" from the remainder of the data. The process is repeated until there are no sideshots left to prune. This iterative process allows for spur traverses with no redundancy to be all identified as sideshots.
This algorithm then removes any benchmarks which were in the control file which had no measurements connected to them.

9.2 2D sideshot identification

The 2D analysis is composed of horizontal distances, horizontal angles, azimuths, and control coordinates. The first three types of data connect stations to one another. A sideshot is defined as a station that is not a control station, is not an occupied station on a horizontal angle, and is only on one distance and angle which are from the same station. If a station has one distance and two angles from the same setup to it, this is not a sideshot as there is angular redundancy to it. A station uniquely located by angle-angle intersection, angle-distance intersection, distance-distance intersection or resection will not be considered as sideshots even though there may be no redundancy to it. This is because the sideshot computation process in EFBP assumes one angle-distance from the same station. Note EFBP automatically recognizes any type of intersection or resection.

The 2D sideshots algorithm looks for stations that are not 2-D control that are only connected to one other station by an angle-distance. These are sideshots

23

and are "pruned" from the remainder of the data. The process is repeated until there are no sideshots left to prune. This iterative process allows for spur traverses with no redundancy to be all identified as sideshots.

Horizontal control that is not connected to any other stations is carried along to the final .xyz file that is imported into a survey/engineering design software system. This is because some horizontal coordinates not connected by the survey can be important in the later computational process. As an example the coordinates may be for a section corner (which was coordinated in a previous survey) which is going to be used in a proportion or subdivision computation.

24

Chapter 10. Estimation of errors in measurements

While the estimation of a survey's random errors is important at all times, in use of least squares it has special meaning. The key item to remember is this is "estimation", and one should not feel there needs to be exactness in the process.

10.1 Error estimation importance in least squares analysis

Least squares minimizes the sum of the weighted residuals squared. A weighted residual is the error estimate divided by its error estimate (thus a snoop number in a .1D or .2D report). That value needs to be squared because a residual can be either positive or negative. Note a weighted residual is unitless as the residual and error estimate have the same units. This enables different types of measurements to be compared to one another as the weighting process makes everything unitless.

Not only do the error estimates enable simultaneous analysis of different measurement types, likewise it enables measurements of the same type to have varying affects (weights) on the final results. A paced distance is a valid form of measurement if assigned a proper error estimate (perhaps 5 ft. per 100 ft.) relative to a EDM distance (.01 ft. plus 5 ppm).

10.2 Error estimation from repetition error plus add-ons

Repetition can be an indicator of an error estimate, but it is usually too small to be used absolutely as an error estimate. As an example, repetition error does not model setup errors at the instrument or prism. It also possible to obtain perfect repetitions, but this does not mean the measurement is perfect. The first form of error estimation EFBP allows is repetition error plus user assigned add-ons which model the errors which repetition cannot model.

10.3 Error estimation without influence of repetition error

Some people feel repetition error should only be used for blunder detection and not in error estimation calculations. A user can thus toggle off error estimation from repetition plus add-ons and instead use user defined constants. If a measurement is not repeated EFBP will use the user defined constants no matter what form of error estimation has been selected.

10.4 Horizontal distance

Horizontal distance error estimation is usually associated with a constant error/add-on plus a ppm (parts per million) error/add-on. The ppm assigns larger error estimates to longer lines. Typical total station error estimate add-ons to repetition error are 0.005-0.01 ft. (0.002-0.004 m) and 2-10 ppm. Typical error estimate constants are 0.01-0.02 ft. (0.004-0.008 m) and 5-20 ppm. Note the constant is in ft. or m. while the ppm is unitless.

10.5 Trigonometric or differential leveling elevation difference

25

Trigonometric elevation difference error estimation is usually associated with a constant error/add-on plus a ppm (parts per million) error/add-on. The ppm assigns larger error estimates to longer lines. It is a well known fact that error in trigonometric leveling propagates faster for longer distances than the error in horizontal distance. Typical total station add-ons to repetition error are 0.02-0.05 ft. (0.008-0.02 m) and 30-100 ppm. EFBP only supports error estimate constants for trigonometric leveling which typically are 0.03-0.10 ft. (0.01-0.03 m). Note the constant is in ft. or m. while the ppm is unitless.

Differential leveling does not usually require any form of repetition. Therefore independent of type of defined error estimation elevation differences at a setup are assigned a user defined error estimate which usually ranges from 0.002-0.01 ft. (0.001-0.003 m.).

10.6 Horizontal angles

Horizontal angle error estimation is usually associated with a constant error/add-on plus a setup error. The setup ensures that shorter lines receive larger error estimates as measuring directions on a shorter line is more difficult than on a longer line. Setup error can be thought of our inability to position exactly over the occupied or sighted station. The error due to setup is the inverse tangent of the setup error (ft. or m.) divided by the length of the line. Typical total station constant error add-on is usually 3-20 seconds, while the constant error estimate is usually 6-30 seconds. Setup error is used in both methods of error estimation, and is generally 0.003-0.01 ft. (0.001-0.003 m.). Setup error is linear units sensitive.

If repetition error plus add-ons is used, error estimate of a angle is:

SQRT (repetition error^2 + constant add-on^2 + BS setup err.^{2} + FS setup err.^{2})

If user defined error estimation is used, error estimate of an angle is:

SQRT (constant err.^{2} + BS setup err.^{2} + FS setup err.^{2})

10.7 Azimuths

EFBP only accepts azimuths in the .ctl file. In the CTL program you are able to assign error estimates to them. The azimuth add-on in the EFBP menu is there for future implementation only.

If the azimuth error estimate has not been entered into the .ctl file, the azimuth error estimate constant in the EFBP menu will be used.

No matter which of these two procedures apply, setup error is always added to the constant error. Setup error is calculated exactly as in horizontal angle error estimation.

The azimuth error estimate is calculated from constant and setup error by:

26

$SQRT\ (constant\ err.^2 + setup\ err.^2)$

10.8 Control coordinates

A control coordinate (horizontal or vertical) can be treated as a measurement with an appropriate error estimate if one desires. This will allow control to adjust along with the rest of one's measurements. The error estimate for control coordinates is in the .ctl file.

Normally one wants to not allow control to adjust and thus when entering control default error estimates are assigned of 0.001 ft. or m. This error estimate is so superior to your other measurements that control will not adjust. Another safeguard to preventing control from adjusting is that if error estimate from user defined constants is selected (do not use repetition error) EFBP will ignore values in .ctl and assign error estimates of 0.001 to all control.

Allowing control to adjust based on non-fixed error estimates has several outstanding abilities. Used with robustness, it is a powerful tool in finding control problems which are often station naming or incorrect data entry. It also lets you evaluate the quality of your measurements without errors in the control coordinates having an affect. It also lets you weight different control accuracies relative to one another.

27

Chapter 11. Validating quality of your measurements in least squares redundancy

Least squares analysis provides a large number of indicators which evaluate the quality of your measurements. The key indicator is the residual which is the difference between the measurement and its adjusted equivalent which is derived from inversing final coordinates. If all of your residuals are within what you would call acceptable random errors in surveying, you should deem the final coordinates acceptable.

11.1 Residual vs. error estimate

A residual and an error estimate for a particular measurement share a very important relationship. Simply looking at a residual does not always give you a clear interpretation without observing its error estimate.

As an example consider two angle residuals of 10 and 60 seconds respectively. At first it looks like the second is much worse than the first and is indicative of a blunder. But the 10 second residual is associated with a 4 second error estimate because of long sight distances, while the 60 second residual has a 12 ft. backsight distance and a 14 ft. foresight distance which created an error estimate (mostly due to setup error) of 80 seconds. Both are acceptable measurements as the residuals and error estimates are within the same reasonable level of magnitude.

One should be concerned when the residual is significantly larger than the error estimate. Simply being larger than the error estimate is not a reason for concern as, from a statistical standpoint, only approximately 67% (one sigma) of our acceptable measurements should have residuals smaller than our error estimates. A general rule of thumb is if any residuals are more than three times the size of their respective error estimates a user is 95% certain there is something wrong with at least one of the measurements or control coordinates. Note this may be numeric (measurement) or a station naming problem. In most cases the problem can be resolved and the data reprocessed without elimination of the measurement.

A significant amount of large residuals of the same sign indicates systematic error. An example is a survey which ties to "good" control that produces all negative distance residuals could be an indicator of an instrument/prism offset constant error.

11.2 Snoop number

Looking at a large number of residuals and error estimates is difficult as one has to mentally make the association of magnitude of the two quantities. To simplify this in both the 1D and 2D least squares reports snoop numbers are associated with all measurements.

A snoop number is the absolute value of the residual divided by the error estimate. If the residual is larger than the error estimate the snoop number is greater than one, and a residual which is smaller than the error estimate

28

produces a snoop number less than one. Snoop numbers greater than three are flagged with asterisks to highlight a potential problem. Usually a series of flagged residuals can be traced to a single problem, and the asterisks go away once the problem is resolved and reprocessing occurs.

The best part of the snoop number concept is how it relates to measurements of the same type which have different error estimates. Let us revisit the example consider two angle residuals of 10 and 60 seconds respectively. At first it looks like the second is much worse than the first and is indicative of a blunder. But the 10 second residual is associated with a 4 second error estimate because of long sight distances, while the 60 second residual has a 12 ft. backsight distance and a 14 ft. foresight distance which created an error estimate (mostly due to setup error) of 80 seconds. The 10 second residual would produce a snoop number of 2.5, and the 60 second residual would produce a snoop number of 0.75. The snoop number shows the 10 second residual indicates a worse observation than the 60 second residual. The 2.5 snoop number is usually regarded as acceptable, but is nearing the concern magnitude and thus may warrant some investigation.

11.3 Root-mean-square error

The root-mean-square (rms) error is associated with a particular type of observation type, and can be thought of as an average residual for that type of observation. To eliminate the affect of the positive/negative nature of residuals, rms error is the square root of the sum of the squares of the residuals divided by the number of that observation type. Note it does not take into consideration the differences in error estimates for a given observation type.

11.4 Root-mean-square snoop number

RMS snoop number is for a given observation type, and is the square root of the sum of the squares of the snoop numbers divided by the number of that observation type. It takes into consideration the differences in error estimates, and is thus a better indicator of data quality than the standard rms error. Note a 2D adjustment may yield rms snoop numbers for horizontal distances and angles of 0.4 and 2.8 respectively. This could be an indicator that your default error estimate parameters for distances should be tightened up and the default error estimate parameters for angles loosened up.

Note that substandard control coordinates which are held fixed will produce higher residuals and rms errors in the measurements. One should be very careful in evaluating your measurement residuals than some of it may be derived from its "fit" to the control coordinates.

11.5 Maximum residual

Maximum residual is the largest (absolute value) difference between measured and adjusted values for a particular type of measurement. One quick way to verify data quality is verifying if the maximum residuals are insignificant in size.

29

Note the maximum residual may not be associated with the largest snoop numbers due to varying error estimates.

11.6 Degrees of freedom

Degrees of freedom is the amount of redundancy in an adjustment. Redundancy is the number of measurements beyond what is needed for unique computation of coordinates. Note this value is computed after sideshots are removed though including them would not change the number of degrees of freedom.

In the 1D adjustment the number of degrees of freedom is the number of benchmarks plus the number of elevation differences minus the total number of stations.

In the 2D adjustment the number of degrees of freedom is the number of control coordinates plus the number of distances plus the number of angles plus the number of azimuths minus the total number of coordinates. Note the number of control coordinates is two times the number of horizontal control stations, and the total number of coordinates is two times the total number of stations in the 2D least squares analysis.

11.7 Standard error of unit weight

The standard error of unit weight is the square root of the sum of the square of the weighted residuals divided by the number of degrees of freedom. A weighted residual is a snoop number. The standard error of unit weight is thus the overall indicator of the fit of the error estimates to the residuals, and should be near one.

11.8 Chi-squared test

The chi-squared test is an analysis of the suitability of the standard error of unit weight. The chi-squared test in EFBP is performed at 95 % (.05 level of significance) confidence in what is termed a two tailed test. The two tailed test means the standard error of unit weight could be too high of low. Obviously a low standard of unit weight (less than one) should be considered positive - you did better than expected - but the chi-squared test could "fail" on this end. Most people would consider doing better than expected not failure but the chi-squared test is simply saying that in the future you may want to start using tighter error estimation parameters.

The chi-squared low and high ends of success/failure are based on the number of degrees of freedom. A lower degree of freedom gives a larger spread. This is because a lower degree of freedom lends itself to more data variability, while a higher number of degrees of freedom means the outliers have less affect on the standard error of unit weight. You are not being punished for higher degrees of freedoms which produces tighter chi-squared high/low tolerances - data variability simply has less affect when you have more degrees of freedom and thus you need tighter high/low tolerances.

30

If the chi-squared test passes you are 95% confident that there is no problem with your data. It is not easy to consistently pass this test as rarely in surveying are you 95% sure about anything. The magnitude of the snoop numbers and residuals should by the judge of suitability even if the chi-squared test fails.

11.9 Minimally constrained vs. constrained adjustment

A minimally constrained adjustment is one where a minimum amount of control is used so that the least squares reports are based solely on one's measurements, and not how one's data "fits" all control that has been tied to. In a 1D minimally constrained least squares one benchmark is held fixed, and in a 2D minimally constrained analysis one control point and one azimuth are held fixed. To derive meaningful results from a minimally constrained adjustment one must ensure a reasonable amount of redundancy can still be achieved in absence of redundant control coordinates.

If sufficient redundancy exists the minimally constrained and constrained analyses can be compared to see if any lack of fit between measurements and control coordinates.

EFBP provides two mechanisms for a quick procedure for obtaining a minimally constrained analysis. If no control (.ctl) file exists EFBP will assume arbitrary 3-D control coordinates of (10000,10000,500) for the first setup and an azimuth of due north to the first sighted station.
If a control file exists one can assign large error estimates to the control and render its affect on the measurement residual statistics null.

31

Chapter 12. Validating repeatability of coordinate production in least squares

Least squares can estimate the quality/repeatability of adjusted coordinates through post-adjustment coordinate standard deviations and error ellipses. EFBP computed these values at a 95% level of confidence.
These computed values are all relative to control coordinate location, i.e., repeatability/reliability of a coordinate close to control is of a smaller magnitude than a coordinate which is a long distance or number of stations from control.

12.1 Introduction - geometry considerations

Geometry of the survey network has an affect on post-adjustment standard deviations and size of error ellipses. It also validates what we know about how error propagates in surveying. As an example, trigonometric leveling would produce larger elevation standard deviations than a differential level survey through the same points. A traverse running north-south will produce smaller northing (Y) than easting standard errors. This is because we measure distances more precisely than angles due to the EDM, and this makes coordinates in the direction of the traverse more reliable than coordinates which are perpendicular to the traverse direction. Finally, an intersection which produces a very non-equilateral triangle will produce higher coordinate standard deviations than an intersection where the triangle is near equilateral.

12.2 F statistic multiplier

To achieve more than one sigma (67%) confidence the F statistic multiplier is applied to all coordinate standard deviations and error ellipse dimensions. The size of the multiplier is a function of desired confidence level (EFBP produces everything at the 95% confidence level) and the number of degrees of freedom. The multiplier decreases in size as the number of degrees of freedom increases.

One standard deviation standard errors and error ellipse dimensions are converted to 95% confidence via the F-statistic multiplier.

This value is (three significant figures):

# of degrees of freedom	F statistic multiplier
1	20.00
2	6.16
3	4.37
4	3.73
5	3.40
6	3.21
7	3.10
8	2.99
9	2.93
10	2.86

# of degrees F statistic
of freedom multiplier

32

11	2.83
12 - 14	2.77
13 - 18	2.71
17 - 26	2.64
25 - 36	2.58
35 - 46	2.55
45 - 60	2.52
59 - 75	2.51
74 - 90	2.50
89 - 120	2.49
119 - 150	2.48
149 - 180	2.47
179 - 210	2.46
210+	2.45

This tells you adding degrees of freedom initially enhances your ability to have better confidence in your work. Note after approximately 25 degrees of freedom the F-statistic goes down very slowly.

This is analogous to why after a certain point repeated measurement of a value does little good in improving its standard deviation in the mean.

The other value affecting standard errors of coordinates and error ellipse dimensions is that the standard error of unit weight is also applied as a multiplier. This makes sense as a standard error of unit weight of 1.0 indicates approximately twice the quality of a standard error of unit of 2.0.

Example:

standard error of unit weight = 1.34
degrees of freedom = 10 -- F- statistic multiplier = 2.86
one standard deviation coordinate error = 0.034 m. (assumes standard error of unit weight = 1.00)

95% confidence coordinate error = 1.34 * 2.86 * 0.034 = 0.13 m.

Weak geometry/ strong geometry in intersections and resections shows up very quickly in evaluations of error ellipses and coordinate errors. Likewise the inherent larger errors in eastings in north-south road projects is evident in reviewing error ellipses.

12.3 Coordinate standard deviations

EFBP produces all post-adjustment coordinate standard deviations at a 95% level of confidence based on the F statistic multiplier and the standard error of unit weight. Coordinate standard errors will be smaller near fixed control as the repeatability of that coordinate is easier than a station which is further away from control.

33

The meaning of the post-adjustment coordinate standard deviation is if you went and performed the same survey over using the same equipment under the same conditions you are 95% sure the second survey's coordinate will be within (plus-or-minus) the standard error about the first survey's coordinate. The standard error gets larger for higher confidence levels.

Post-adjustment standard deviations are very much a function of survey geometry. As an example, a north-south traverse will generally produce smaller northing errors than the easting error for the same point. The easting errors could be reduced by astronomic observations, additional control, or ties in an east-west direction.

12.4 Error ellipses

Error ellipses are output by EFBP at 95% confidence, and are thus multiplied by the F-statistic multiplier and the standard error of unit weight. An error ellipse is defined by SU - semi-major axis, SV - semi--minor axis, and T - angle of the semi-major axis off north (clockwise positive). A semi-axis is from the center to the external edge of the ellipse. The semi-major is the longest axis of the ellipse, and the semi-minor is the shortest axis and is 90 degrees from the semi-major axis. The least squares adjusted coordinate is at the center of the ellipse. Error ellipses will be smaller near fixed control as the repeatability of that coordinate is easier than a station which is further away from control.

The meaning of the error ellipse is if you went and performed the same survey over using the same equipment under the same conditions you are 95% sure the second survey's coordinate will be within (plus-or-minus) the error ellipse about the first survey's coordinate. The error ellipse gets larger for higher confidence levels.

Error ellipses are very much a function of survey geometry. As an example consider a tower a long distance from the job site which is being used simply as a direction check from a number of stations. Since it is very doubtful good geometry of equilateral triangles exists, the error ellipse for the intersection will be large, especially in the direction of the survey lines to the tower. In this case this is expected, and there is nothing wrong with the measurements to the tower unless large residuals exist to it. It still provides a good directional check for the job, but its final coordinates should not be treated as fixed if another survey ties to it.

12.5 Repeatability of derived quantities

The post-adjustment statistics of coordinate standard deviation and error ellipse indicators of reproducibility if one performed the same survey over under the same conditions. A subsequent survey which does not exactly follow this rule should not be using this information as you not comparing relatable items. Any post-adjustment standard deviation or error ellipse should definitely not be regarded as the error in the "absolute position" of the point as that in no way follows the rules which they are based on. One should be extremely careful in understanding the limits of the interpretation of error ellipses.

34

Chapter 13. Theory of least squares solution

While the theory of least squares adjustment as it applies to surveying can be found in well documented form in several text books, several of the important concepts are presented here in simple form.

13.1 Minimization

Least squares minimizes the sum of the squares of the weighted residuals. A weighted residual is the residual divided by the error estimate. That quantity is what is squared, and each measurement (observation) needs to included in the summation.

To obtain a minimum the sum of the squares of the weighted residuals are subjected to partial differentiation with respect to the parameter which is the residual, and that equation is set equal to zero. Since it is desired to solve for the unknown coordinates, the observation equation is used to substitute for the residual in terms of the unknowns. An observation equation defines a measurement plus its residual in terms of an equation which defines the measurement in terms of coordinates.

The observation equation for an elevation difference is simply the "to" station's elevation minus the "from" station's elevation. The observation equation for a horizontal distance is the pythagoreum theorem "inverse" of the coordinates. The observation equation for an azimuth is the tangent inverse of the change in eastings divided by the change in northings. The observation equation for a horizontal angle is the difference between the foresight and backsight directions, and is thus the similar to an azimuth applied twice where the two values are differenced. The observation equation for a control coordinate is simply the input coordinate is equal to its adjusted value plus the residual.

The second derivative can be taken and solved. This results in a positive value which assures we have computed a minimum (a negative value assures a maximum has been calculated).

13.2 Linearization

Certain observation equations cannot be directly solved because they are non-linear. A non-linear equation is any equation with any exponentials besides one (including square root) or any trigonometric functions which include unknown coordinates. The observation equations for differential leveling and control coordinates are linear, and the observation equations for distance (square root and squared), azimuths (tangent inverse), and angles (tangent inverse) are non-linear.

The 1D least squares adjustment is thus linear and is solved directly. Solved directly means the elevations are directly solved for.

35

The 2D least squares adjustment contains non-linear equations and thus requires linearization. Linearization is performed using a Taylor's series expansion where all but the first order differentials are considered negligible. This approach requires input of approximate values for all unknowns (coordinates), and the solution is actually for updates to the approximate coordinates. The update process is iterative (note the 1D least squares does not iterate) where the updates to the unknowns eventually become insignificant. In cases of large blunders the solution may actually get worse as iterations proceed (divergence). EFBP uses 0.001 ft. or m. as the maximum update convergence criteria and quits if divergence or 10 iterations occur.

Both minimization and linearization involve calculus (differentiation). Note a user of least squares does not have to understand the derivation and is thus not required to have knowledge of calculus.

13.3 Normal equations

Least squares forms and solves a "n x n" system of equations where n is the number of unknowns. The number of unknowns is the number of stations in a 1D adjustment (one elevation per station) and in the 2D adjustment it is two times the number of stations (two coordinates per station). The equations which are formed and solved are called the normal equations.

13.4 Cholesky solution (positive definite systems of equations)

System of equations can be solved by a variety of methods. Least squares normal equations in surveying always belong to a class of equations which are positive-definite. The understanding of the positive definite class of equations is a topic of linear algebra. A user of least squares does not need to know linear algebra.

The positive-definite classification can be taken advantage of and solved using the Cholesky (square root) solution. This type of solution is less prone to round-off and is significantly faster than more generic forms of solution. EFBP uses the Cholesky solution in the 1D and 2D least squares algorithms.

The normal equations in surveying are always symmetric (the term in row 2, column 5 equals the term in row 5, column 2). This allows EFBP to only have to store approximately 50% of the terms which saves on computer storage and the amount of necessary computations.

13.5 Variance-covariance matrix

If the normal equations are written as NX=C where N is a "n x n" system of coefficients, X is an "n x 1" vector of unknowns, and C is the "n x 1" vector of constants, let N^{-1} be another "n x n" system of coefficients which results from NX=C being re-written as $X=N^{-1}C$.

N^{-1} is called the inverse of N, and it can also be shown that it represents the variance-covariances of the unknown coordinates. N^{-1} is thus the coefficients from which post-adjustment coordinate standard deviations and error ellipses can

36

be derived. To obtain this information only a very small fraction of terms in N^{-1} (2 * n terms) need to be calculated. The Cholesky solution can again be used to efficiently calculate these variance-covariance terms.

If no redundancy existed, simple error propagation of coordinate equations and variance-covariance would yield identical results. The simply equation based error propagation cannot efficiently handle using redundancy in a survey in calculating the results of coordinate standard deviations and error ellipses.

A user does not have to understand the derivation of the variance-covariance matrix. A user simply needs to know how post-adjustment error analysis should be utilized.

13.6 Sparsity of the normal equations

Solving a system of "n equations, n unknowns" where n is large is a very time consuming problem even on a fast computer. It is compounded by the fact that in the 2D adjustment the system has to be solved multiple times as the solution iterates to convergency.

The normal equations in survey adjustments tend to be sparse, i.e., many of the coefficients are zero. This is true because a non-zero term indicates two stations are directly connected by a measurement. While it is possible that every station in a survey is directly connected to every other station by a measurement (no zero terms), this is highly unlikely.
In survey networks a particular station is usually only directly connected by measurements to a small subset of the total number of redundant stations. The number of zero terms is thus quite large as a percentage of the total number of terms.

What you want to avoid is having the computer is operate on zero terms - adding zero to a number, or multiplying zero times a number and adding it to another number, are unnecessary operations. It is also possible to not store zero terms in a computer register. The location of the zero terms in the normal equations defines if an algorithm can eliminate storage of zero terms and eliminate most of the addition of zero terms.

13.7 Taking advantage of sparsity - bandwidth optimization

EFBP uses a bandwidth optimization process in taking advantage of the sparsity of the normal equations. This process is actually a station reordering which places the zero terms in a grouped area so that the algorithm knows not to store or operate on those terms. The bandwidth optimization process places stations which are directly connected by a measurement(s) in close proximity in the reordered list.

While the number of bandwidth optimization algorithms is immense, EFBP uses a simple one which works very well for survey network type station connectivity. Note it applies to the survey after sideshots have been eliminated. The algorithm starts at a station with the most number of connected stations (connected by a measurement). Connected stations are added to the list. Next stations not in the list are added which are connected to the second station in the list, then the third station, etc. Eventually every station is in the list.

The term maximum bandwidth refers to the reordered station list. The furthest "distance" in the list between two stations directly connected by a measurement is the maximum bandwidth. This is computed as the list is being built.

The bandwidth is displayed as it is one of the elements in estimating how long the least squares will take to process. Survey networks with small redundancy will tend to have a smaller bandwidth as a percent of the number of stations in the network. As a survey becomes more interconnected (redundant) the bandwidth becomes a larger percent of the total number of stations.

Bandwidth optimization is essential for efficient solution of least squares problems. Knowledge of how the bandwidth procedure works is not required for a user of EFBP.

38